Non-Response in Dynamic Panel Data Models

Cheti Nicoletti
ISER, University of Essex, Wivenhoe Park
Colchester CO4 3SQ, United Kingdom
E-mail: iserg016@essex.ac.uk

Abstract

This paper stresses the links that exist between concepts that are used in the theory of model reduction and concepts that arise in the missing data literature. This connection motivates the extension of the *missing at random* (MAR) and the *missing completely at random* (MCAR) concepts from a static setting, as introduced by Rubin (1976), to the case of dynamic panel data models.

Using this extension of the MAR and MCAR definitions, we emphasize the limits of some tests and procedures, proposed by Little (1988), Diggle (1989), Park and Davis (1993), Taris (1996) and others, to verify the ignorability of the missing data mechanism.

Key words: dynamic panel model, attrition, non-response, missing at random, missing completely at random, statistical model reduction.

JEL classification: C33, C34

1. Introduction

The non-response problem is ignorable for a regression model of interest if we can make inference on this model ignoring the process that causes missing data. In other words, ignorability requires that the inference on the model of interest, neglecting the missing data generating mechanism, [1] be affected neither in terms of distortion nor efficiency. The conditions that allow one to neglect the selection process are given in Rubin (1976) and Little and Rubin (1987) for the cross-sectional case. In particular, these authors introduced the concepts of *missing at random* (MAR), *observed at random* (OAR), *missing completely at random* (MCAR) and *parameter distinctness*.

The extension of MAR and MCAR to the panel data case is straightforward when the data are independently and identically distributed across units and over time. However, while the data on different units can generally be assumed to be independent, at least conditionally on some exogenous variables, the repeated observations on the same unit are likely to be dependent. This is the reason for the widespread use of dynamic regression models. In this paper, we derive the set of

Part of this paper is based on work carried out during a visit to the European Centre for Analysis in the Social Sciences (ECASS) at the Institute for Social and Economic Research, University of Essex supported by the Access to Research Infrastructure action under the EU Improving Human Potential Programme. Special thanks are given to Pierre Hoonhout and Franco Peracchi for their comments on this paper.

[1] Henceforth we will call the 'missing data generating mechanism' more briefly 'missing data process' or 'selection process' (or 'mechanism').

conditions under which the selection process can be safely ignored when making inference on a dynamic regression model. Our approach is triggered by ideas that arise in the theory of statistical models reduction (see Engle, Hendry and Richard 1983; Hendry 1995).

Using the definition of MAR and MCAR for panel data, we outline the limits of some tests proposed in the literature to verify the MCAR in multivariate data, as the tests and the procedures presented in Little (1988), Diggle (1989), Park and Davis (1993), Taris (1996) and some of the variable addition tests, presented in Verbeek and Nijman (1992).

The paper is organized as follows: in Section 2, we give the formal definitions of MAR and MCAR; in Section 3, we emphasize the limits of some tests for MCAR and MAR for multivariate data; and in Section 4, we give some conclusions.

2. Definitions of MAR and MCAR

In this section, after some preliminary definitions and general notation given in Section 2.1, we define the conditions of MAR and MCAR. These conditions must be properly redefined for different types of models of interest. For this reason, we dedicate separate sections to define MAR and MCAR for different types of model: Section 2.2 for marginal models; Section 2.4, for conditional models; Section 2.5 for dynamic panel models with general response patterns; Section 2.6 for dynamic panel models with attrition; and Section 2.7 for dynamic panel models with explanatory variables.

Furthermore, we emphasize the differences between our definitions and those given by other authors. In particular, for the cross-sectional data case, we consider the definitions of Rubin (1976), Little and Rubin (1987) and Heitjan and Rubin (1991) (see Section 2.3); whereas, for the multivariate data case, we examine the definitions given by Robins and various co-authors (see Section 2.8).

Finally, in Section 2.9, we conclude by describing some further possible extensions of the MAR and MCAR concepts.

2.1 General statement and notation

We begin by considering the cross-sectional data case and focus our attention on a model for the variable y, $\{\mathbf{Y}, f(y;\theta), \theta\in\Theta\}$; where $\mathbf{Y}$ is the sample space, $f(y;\theta)$ is a family of probability distributions indexed by θ, a vector of parameters of interest, and Θ is the parameter space. The variable, y, is missing if the dummy variable $r=0$, and observable if $r=1$. Let us indicate with y^m the missing variable associated with $r=0$ and y^o, the observed variable associated with $r=1$. By analogy, let $\mathbf{Y}^m$ and $\mathbf{Y}^o$ be the subspaces of $\mathbf{Y}$ for the missing and observed variables, respectively. Let $\{\mathbf{Y}\times\mathbf{X}, f(r,y;\varphi)\ , \varphi\in\Psi\}$ be the joint model for (r,y). Finally, let $f(r|y;\phi)$ be the probability that $r=1$ or 0, conditional on the variable y, that is, the selection mechanism or the missing data process, where ϕ is a vector of nuisance parameters.

We define three different types of likelihood functions that we could use to make inference on the model of interest in the presence of missing data. We write the likelihood function for a single observation, but the extension to a random sample of N units is straightforward.

The first likelihood,

$$L_T = \left(f(y^o;\theta)\right)^r, \qquad (1)$$

let's say the truncated likelihood, does not take account of the missing data in the variables, as it considers only the truncated sample of observable values.

The second likelihood function,

$$L_c = \left(f(y^o;\theta)\right)^r \left(\int_{\mathbf{Y}^m} f(y^m;\theta)dy^m \right)^{1-r} = \int_{\mathbf{Y}^m} f(y;\theta)dy^m \ , \tag{2}$$

let's say the censored likelihood, considers both observed and unobserved variables, but not the missing data process.

Finally, for the third likelihood function,

$$L_I = \int_{Y^m} f(y;\theta) f(r \mid y;\phi)dy^m \ , \tag{3}$$

let's say the likelihood with informative missing data, the model of interest and the selection mechanism are considered jointly and the missing variables are 'integrated out'.

In the following, we say that the selection mechanism is *weakly ignorable* if we can make a correct and efficient inference based on the likelihood (1) or (2) disregarding the selection process. Whereas we say that the selection mechanism is *strongly ignorable* if any type of inference can be made correctly and efficiently without considering the selection process.[2]

2.2 Definitions of MAR and MCAR for a marginal model of interest

Following Heitjan and Rubin (1991), likelihood-based inference on θ can be made ignoring the data mechanism if:

1. $f(y,r;\varphi)$ factorizes in $f(y;\theta)$ $f(r|y;\phi)$, where θ e ϕ are variation free or, as Rubin (1976) says: […the parameter ϕ is distinct from θ], that is […their joint parameter space factorizes into a θ-space and a ϕ-space],
2. y is missing at random (MAR); that is $f(r \mid y^m;\phi)$ takes the same value for any y^m belonging to the space of possible missing values, say, $\mathbf{Y}^m$, that is a subspace or the entire sample space of y, $\mathbf{Y}$.

When conditions (1) and (2) are satisfied, we say that the missing data mechanism is weakly ignorable or, more briefly, ignorable. Moreover, we say that the missing data mechanism is strongly ignorable if, besides (1) and (2), the following condition is satisfied:

3. y is observed at random (OAR), that is $f(r \mid y^o, y^m;\phi)$ takes the same value for any y^o belonging to the space $\mathbf{Y}$.

In accordance with the theory of model reduction, we call 1 the statistical cut assumption. When the conditions 2 and 3 are satisfied, then y and r are independent and we will denote this independence as $y \perp r$. Conditions 2 and 3 together constitute MCAR. Obviously, it is implicitly assumed that the model of interest, $f(y;\theta)$, is the reduced model resulting from an admissible reduction of the data generating process.

To give some indication as to whether weak ignorability indeed suffices for a correct likelihood-based inference when the selection process is disregarded,[3] we note that the likelihood ratio when disregarding the selection process is equal to the likelihood ratio when taking it into account

$$\frac{L_I(\theta_0)}{L_I(\theta_1)} = \frac{\int_{Y^m} f(y;\theta_0) f(r \mid y;\phi)dy^m}{\int_{Y^m} f(y;\theta_1) f(r \mid y;\phi)dy^m} = \frac{\int_{Y^m} f(y;\theta_0)dy^m}{\int_{Y^m} f(y;\theta_1)dy^m} = \frac{L_c(\theta_0)}{L_c(\theta_1)} .$$

[2] The definition of strong ignorability used in this paper coincides with Verbeek and Nijman (1992)'s definition, whilst the definition of (weak) ignorability is not equivalent to their definition.

[3] For a more formal proof, see Rubin (1976).

The observed data allow the identification of the probability distribution $f(y|r=1)$, which is not equivalent to the marginal distribution of y^o, $\int_{\mathbf{Y}^m} f(y;\theta)dy^m$. To ensure that inferences based on $f(y|r=1)$ and $\int_{\mathbf{Y}^m} f(y;\theta)dy^m$ be equivalent, the data must be MCAR and the variation-free condition must be satisfied. Indeed, under these conditions, the following equality is true:

$$\int_{\mathbf{Y}^m} f(y;\theta)dy^m = \int_{Y^m} \frac{f(y;\theta)\, f(r\mid y;\phi)}{\int f(y;\theta)\, f(r\mid y;\phi)dy} dy^m .$$

2.3 Differences among MAR definitions

The definition of MAR given here differs slightly from the definition given in Little and Rubin (1987). Whilst we require that the selection mechanism be constant only when y^m belongs to the subspace of possible missing values, $\mathbf{Y}^m \subset \mathbf{Y}$, Little and Rubin require that the probability of observing *y* be constant for any y^m belonging to **Y**. Our definition of MAR is equivalent to the enlarged definition of coarsened at random given by Heitjan and Rubin (1991), where the definition of MAR is extended to any type of coarsened data (censored, heaped, grouped, rounded, etc.). We present this extension of the concept of MAR in Appendix A.

Whilst Little and Rubin (1987) define MAR as the condition which ensures a correct inference based on the truncated likelihood, we define MAR in the same way as Heitjan and Rubin (1991); i.e., as the condition which allows a correct inference based on the censored likelihood. When the censored and truncated likelihood functions are equal, the two definitions coincide. In particular, this is true when $\mathbf{Y}^m = \mathbf{Y}$.

If the selection process is deterministic, that is if the dummy variable *r* conditioning on *y* is degenerate, then we say that the data are MAR; in contrast, Little and Rubin (1987) say that the data are not MAR in this case. This distinction may lead to confusion, the most notable example of which is the case of a censored variable for which no values are observed when the variable belongs to a specific subset, $\mathbf{Y}^m \subset \mathbf{Y}$. This is indeed an instance in which correct inference can be based on the censored likelihood, and the censored and truncated likelihood functions are not equal. The latter observation is proved in Appendix B.

This observation holds more generally. Suppose we can divide the sample space into *s* disjoint subspaces, $\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_s$, and suppose for every missing variable we know to which subspace it belongs; moreover, assume that the selection process is such that $\Pr(r=1 \mid y \in \mathbf{Y}_j) = c_j$, where c_j is constant within the same subspace; then we can say that the data are MAR and that inference can be based on the censored likelihood.

2.4 MAR and MCAR for a conditional model of interest

As remarked by Shih (1992), some authors do not explicitly mention the variation-free condition (the condition 1 in Section 2.2). This condition is often implicitly assumed to be valid in econometric literature; in particular, econometricians usually implicitly assume that the conditional or marginal model of interest is the result of an admissible reduction of the data generating process.

In this section, to avoid any misunderstanding, we explicitly state all the conditions necessary to ignore the selection mechanism when the model of interest is a conditional one.

Let us assume that we are interested in the conditional model for the variable y, given a set of variables x belonging to the space X, {**Y**, $f(y|x;\theta)$, $\theta\in\Theta$}, where **Y** is the sample space, $f(y|x;\theta)$ is a family of conditional probability distributions indexed by the parameter θ, and Θ is the parameter space. Furthermore, let us assume that the true data generating process is the joint model {**Y**×**X**x**R,** $f(y,x,r;\varphi)$, $\varphi\in\Phi$}. Then, to make a likelihood-based inference on the conditional model of interest neglecting the selection process, that is the model {**R,** $f(r|\mathrm{y,x};\gamma)$, $\gamma\in\Gamma$}, the following conditions must be satisfied:

1. the following two statistical cuts must be satisfied

   $f(y,x,r;\varphi) = f(y,r|x;\psi)\, f(x;\phi)$, and,

   $f(y,r|x;\psi) = f(y|x;\psi_1)\, f(r|y,x;\psi_2)$;
2. the independence of r from y, given x, to ensure the MCAR condition; the independence of r from y^m given x to ensure the MAR condition.

Again, we say that the selection mechanism is weakly ignorable if condition 1 and MAR are satisfied, while we say that the selection mechanism is strongly ignorable if condition 1 and the MCAR are satisfied.

2.5 MAR and MCAR for a dynamic panel data model

Panel data are constituted by a sample of units followed over time and they are often used to estimate dynamic models. Dynamic models are those in which the dependent variable is explained by its past and/or the present and past of other variables. In the following, we will consider a generic panel composed of N units followed for T consecutive waves.

As already mentioned, in the case of a random sample of N units observed at a single occasion (T=1), the definitions of MAR and MCAR stated in Section 2.2 apply. Indeed, (y_i, r_i) are identically and independently distributed (i.i.d.), and the joint likelihood factorizes into the product of N identical likelihood, $f(y_1,...,y_N,r_1,...,r_N;\varphi) = \prod_{I=1}^{N} f(y_i,r_i;\varphi)$. This is no longer true when the variables observed at consecutive time periods, for a specific unit, are not independent.

The definition of weak and strong ignorability can be easily extended to the case of a panel, considering a joint model for $\mathbf{y}_{i,1}^{T}$. Condition 1 in Section 2.2 is substituted by a condition of initial cut:

1'. $f(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T};\varphi) = f(\mathbf{r}_{i,1}^{T}|\mathbf{y}_{i,1}^{T};\phi)\, f(\mathbf{y}_{i1}^{T};\theta)$,

where $\mathbf{y}_{i,1}^{T}$ is the vector of the variables $y_{i,t}$ for the i-th unit and for $t=1,\ldots,T$, while $\mathbf{r}_{i,1}^{T}$ is the vector associated with the response pattern of the i-th unit, that is the vector of the dummies $r_{i,t}$, taking value *1* when the variable $y_{i,t}$ is observed, and *0* otherwise.

Conditions 2 and 3 are replaced by the equivalent assumptions:

2'. $f(\mathbf{r}_{i,1}^{T}|\mathbf{y}_{i}^{o}, \mathbf{y}_{i}^{m};\phi) = f(\mathbf{r}_{i,1}^{T}|\mathbf{y}_{i}^{o};\phi)$,

3'. $f(\mathbf{r}_{i,1}^{T}|\mathbf{y}_{i}^{o}, \mathbf{y}_{i}^{m};\phi) = f(\mathbf{r}_{i,1}^{T};\phi)$,

where $\mathbf{y}_{i,1}^{m,t}$ is the sub-vector of the missing variables and $\mathbf{y}_{i,1}^{o,t}$ is the one of observable variables of the vector $\mathbf{y}_{i,1}^{t}$.

The variables observed for a unit are likely to be dependent from their past; that is, the factorization $f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right) = \prod_{t=1}^{T} f\left(y_{i,t}, r_{i,t}; \varphi\right)$ is not valid and we have to use the sequential factorization $f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right) = \prod_{t=1}^{T} f\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}; \varphi\right)$.[4] In other words we assume that $\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}\right)$ be identically and independently distributed across units and time. In this case, a more appropriate model of interest is a dynamic one, which tries to explain *y* as a function of its past, $f\left(y_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}; \theta\right)$. Then it is useful to restate the conditions 1', 2' and 3' in terms of sequential models.

Condition 1' requires that:

a1. the sequential cut,

$$\prod_{t} f\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}; \varphi\right) = \prod_{t} f\left(y_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}; \theta\right) \prod_{t} f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}; \phi\right),$$

must be applicable;

a2. *r* does not Granger cause *y*, that is,

$$f\left(y_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}; \theta\right) = f\left(y_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}; \theta\right).$$

Further conditions that 2' and 3' require are:

b. $f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}; \phi\right) = f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}; \phi\right)$ or $r_{i,t} \perp \mathbf{y}_{i,1}^{t} \middle| \mathbf{r}_{i,1}^{t-1}; \phi$.

The condition b can be broken down into two parts:

b1. $r_{i,t} \perp \mathbf{y}_{i,1}^{m,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,t}; \phi$,

b2. $r_{i,t} \perp \mathbf{y}_{i,1}^{o,t} \middle| \mathbf{r}_{i,1}^{t-1}; \phi$.

In the case of dynamic panel data, b1 is the sequential MAR condition, b2 is the sequential OAR condition, while b is the sequential MCAR assumption. The conditions a1, a2 and b1 ensure that the missing data mechanism is weakly ignorable for the maximum likelihood estimation of $f\left(y_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}; \theta\right)$, while the conditions a1, a2, b1 and b2 ensure strong ignorability in any inference.

If we consider a maximum likelihood that completely eliminates the units for which there is a wave non-response, the weak ignorability is no longer a sufficient condition and we need the MCAR condition, as for any other type of inference (such as the sampling distribution inference).

2.6 MAR and MCAR conditions in a dynamic panel model with attrition

In this section, we present a proposition which gives a set of necessary and sufficient conditions for the weak ignorability of the selection mechanism; that is, for the conditions 1' and 2', in the case of attrition.

[4] To simplify notation in the sequential models, we implicitly condition on the set of initial conditions.

Proposition *Let* $\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}\right)$ *be i.i.d. across units and time, and let* $f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \phi\right) = \prod_{t=1}^{T} f\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}; \phi\right)$ *be the associated data generating process. Let* $y_{i,t}$ *be observed when* $r_{i,t}$ *takes value 1, and missing when* $r_{i,t}=0$. *Further, whenever* $r_{i,t}=0$, *let* $r_{i,s}=0$ *for any s>t.*

Then, if the condition a2 (r does not Granger cause y) is true, a set of necessary and sufficient conditions for the weak ignorability of the selection mechanism is:

a1. it must be possible to operate a sequential cut

$$\prod_t f\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}; \varphi\right) = \prod_t f\left(y_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}; \theta\right) \prod_t f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}; \phi\right),$$

c1. $r_{i,t} \perp y_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}$.

Proof

First, we prove that a1 and c1 are sufficient conditions to ensure 1' and 2', that is, weak ignorability.

Applying the condition of Granger non-causality to the factorization a1, we obtain:

$$f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right) = \prod_t f\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}; \varphi\right) =$$

$$= \prod_t f\left(y_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}; \theta\right) \prod_t f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}; \phi\right) = f\left(\mathbf{y}_{i1}^{T}; \theta\right) f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{y}_{i,1}^{T}; \phi\right),$$

so that a1 and a2 ensure the initial cut, 1'.

Let us assume that a unit, i, drops out at d-th wave, and let us rewrite the model as the product of three factors:

$$f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right) = L_1 \cdot L_2 \cdot L_3,$$

where

$$L_1 = \left[\prod_{t=1}^{d-1} f\left(y_{i,t}^{o} \middle| \mathbf{y}_{i,1}^{o,t-1}; \theta\right) \prod_{t=2}^{d-1} f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,t}; \phi\right)\right],$$

$$L_2 = \left[f\left(y_{i,d}^{m} \middle| \mathbf{y}_{i,1}^{o,d-1}; \theta\right) \quad f\left(r_{i,d} \middle| \mathbf{r}_{i,1}^{d-1}, \mathbf{y}_{i,1}^{o,d-1}, \mathbf{y}_{i,d}^{m}; \phi\right)\right],$$

$$L_3 = \left[\prod_{t=d+1}^{T} f\left(y_{i,t}^{m} \middle| \mathbf{y}_{i,1}^{o,d-1}, \mathbf{y}_{i,d}^{m,t-1}; \theta\right) f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,d-1}, \mathbf{y}_{i,d}^{m,t}; \phi\right)\right].$$

In a likelihood-based inference on the parameter θ, we *must eliminate the unobserved variables through the integration from the likelihood,* $f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right)$, *in the following way:*

$$\int f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}; \varphi\right) d\mathbf{y}_{i,1}^{m,T} = \int L_1 \cdot L_2 \cdot L_3 \, d\mathbf{y}_{i,1}^{m,T}.$$

The factor L_1 *does not depend on unobserved variables, so it can be taken out of the integral sign. Since we have assumed that* $\left(y_{i,t}, r_{i,t} \middle| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}\right)$ *are i.i.d., and that* $f\left(r_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}; \phi\right)$ *has the same distribution form for each t, then the condition c1,* $r_{i,t} \perp \mathbf{y}_{i,t} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}$, *is equivalent to* $r_{i,t} \perp \mathbf{y}_{i,t}^{m} \middle| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,t-1}$, *so that the factor,* $f\left(r_{i,d} \middle| \mathbf{r}_{i,1}^{d-1}, \mathbf{y}_{i,1}^{o,d-1}, y_{i,d}^{m}; \phi\right) = f\left(r_{i,d} \middle| \mathbf{r}_{i,1}^{d-1}, \mathbf{y}_{i,1}^{o,d-1}; \phi\right)$, *can be taken out of the integral sign too.*

For any t>d, $\left(r_{i,t}\middle|r_{i,d}=0\right)$ *is independent of any variable because if* $r_{i,d}$ *= 0, then* $\Pr(r_{i,t}=0|r_{i,d}=0)=1$ *and* $r_{i,t}$ *becomes degenerate. If* $r_{i,t}$ *=0, then* $f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{o,d-1},\mathbf{y}_{i,d}^{m,t};\phi\right)=1$, *consequently the selection mechanism,* $f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{o,d-1},\mathbf{y}_{i,d}^{m,t};\phi\right)$, *cancels out of the likelihood for any t>d.*

The integrated likelihood becomes:

$$\int f\left(\mathbf{y}_{i,1}^{T},\mathbf{r}_{i,1}^{T};\varphi\right)d\mathbf{y}_{i,1}^{m,T}=L_1\cdot f\left(r_{i,d}\middle|\mathbf{r}_{i,1}^{d-1},\mathbf{y}_{i,1}^{o,d-1};\phi\right)\cdot\int\prod_{t=d}^{T}f\left(y_{i,t}^{m}\middle|\mathbf{y}_{i,1}^{o,d-1},\mathbf{y}_{i,d}^{m,t-1};\theta\right)d\mathbf{y}_{i,d}^{m,T}.$$

Since $\int\prod_{t=d}^{T}f\left(y_{i,t}^{m}\middle|\mathbf{y}_{i,1}^{o,d-1},\mathbf{y}_{i,d}^{m,t-1};\theta\right)d\mathbf{y}_{i,d}^{m,T}=1$, *we can rewrite this as:*

$$\left[\prod_{t=1}^{d-1}f\left(y_{i,t}^{o}\middle|\mathbf{y}_{i,1}^{o,t-1};\theta\right)\right]\cdot\left[\prod_{t=2}^{d-1}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{o,t};\phi\right)\right]\ f\left(r_{i,d}\middle|\mathbf{r}_{i,1}^{d-1},\mathbf{y}_{i,1}^{o,d-1};\phi\right).$$

Given that θ *e* ϕ *are variation free, we can make inference on the parameter* θ *ignoring the selection mechanism, that is considering the likelihood for the observable variables:* $\prod_{t=1}^{d-1}f\left(y_{i,t}^{o}\middle|\mathbf{y}_{i,1}^{o,t-1};\theta\right)$.

In this way, we have also proved that the condition 2' is true:

$$f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i}^{o},\mathbf{y}_{i}^{m};\phi\right)=\prod_{t=1}^{T}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t};\phi\right)=\prod_{t=1}^{T}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t-1};\phi\right)=f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i}^{o};\phi\right)$$

In the following, we prove that a1 and c1 are necessary conditions to ensure 1' and 2'. We begin by proving that when the initial cut 1' operates and condition a2 holds, then a1 is true.

Using condition 1', we can state that:

$$f\left(\mathbf{y}_{i,1}^{T},\mathbf{r}_{i,1}^{T};\varphi\right)=f\left(\mathbf{y}_{i1}^{T};\theta\right)f\left(\mathbf{r}_{i1}^{T}\middle|\mathbf{y}_{i,1}^{T};\phi\right)=\prod_{t}f\left(y_{i,t}\middle|\mathbf{y}_{i,1}^{t-1};\theta\right)\prod_{t}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{T};\phi\right).$$

Since condition a2 may be restated as $r_{i,t}\perp y_{i,t+1}^{T}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t}$, [5] *we can rewrite the joint likelihood as:*

$$f\left(\mathbf{y}_{i,1}^{T},\mathbf{r}_{i,1}^{T};\varphi\right)=\prod_{t}f\left(y_{i,t}\middle|\mathbf{y}_{i,1}^{t-1};\theta\right)\prod_{t}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t};\phi\right),$$

so that $f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i,1}^{T};\phi\right)=\prod_{t}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t};\phi\right)$ *and the sequential cut a1 operate.*

The equality, $f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i,1}^{T};\phi\right)=\prod_{t}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t};\phi\right)$, *and condition 2' imply that:*

$$f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i,1}^{T};\phi\right)=\prod_{t}f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{o,t}\phi\right).$$

$\left(y_{i,t},r_{i,t}\middle|\mathbf{y}_{i,1}^{t-1},\mathbf{r}_{i,1}^{t-1}\right)$ *are i.i.d. across units and time; hence* $f\left(r_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{o,t}\phi\right)$ *maintains a common form for any t. Since for t>d,* $\left(r_{i,t}\middle|r_{i,d}=0\right)$ *is a degenerate variable independent of the past value of y, and for t=d, the sequential selection model does not depend on the value of y at time t, the last equality prove that c1 is satisfied.*

[5] For a proof of this last equivalence, see Florens and Mouchart (1982).

The theorem states that, in the case of dynamic panel data with attrition, the condition y does not Granger cause, r, $r_{i,t} \perp y_{i,1}^{t-1} \big| \mathbf{r}_{i,1}^{t-1}$, is neither necessary nor sufficient condition for the MAR assumption. This Granger non-causality is instead a necessary but not sufficient condition for MCAR. The theorem also proves that the sequential MAR condition is given by (c1) $r_{i,t} \perp y_{i,t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}$, in the case of the problem of attrition. In other words, in the case of attrition, the conditions (a1), (b1) and (c1) ensure a correct likelihood-based inference on the dynamic model of interest, i.e. the weak ignorability.

It is easy to prove that the strong ignorability for a dynamic panel model with attrition requires the sequential MCAR condition, $r_{i,t} \perp \mathbf{y}_{i,1}^{t} \big| \mathbf{r}_{i,1}^{t-1}$, instead of the sequential MAR one.

2.7 MAR and MCAR conditions in a dynamic panel model with explanatory variables

The definitions of MAR and MCAR can be easily modified to cover conditional models of the form, $f(y_{i,t} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}; \theta)$, where explanatory variables x are added to the dynamic panel model.

Let $f(y_{i,t} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}; \theta)$ be the model of interest, let $f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T}, \mathbf{x}_{i,1}^{T}; \varphi\right) = \prod_t f\left(y_{i,t}, r_{i,t}, x_{i,t} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t-1}; \varphi\right)$ be the associated data generating process and let the missing data problem be narrowed down to the attrition problem; then, it is easy to prove that weak ignorability requires the following conditions:

d1. the weak exogeneity of x, that is

$$\prod_t f\left(y_{i,t}, r_{i,t}, x_{i,t} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t-1}; \varphi\right) =$$
$$= \prod_t f\left(y_{i,t}, r_{i,1} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t}; \varphi_1\right) \prod_t f\left(x_{i,t} \big| \mathbf{x}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}; \varphi_2\right),$$

d2. the sequential cut

$$\prod_t f\left(y_{i,t}, r_{i,t} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t}; \varphi_1\right) =$$
$$= \prod_t f\left(y_{i,t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t}; \theta\right) \prod_t f\left(r_{i,t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t}, \mathbf{x}_{i,1}^{t}; \phi\right),$$

d3. the Granger non-causality

$$y_{i,t} \perp \mathbf{r}_{i,1}^{t-1} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t},$$

d4. the sequential MAR condition

$$r_{i,t} \perp y_{i,t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t}.$$

In the case of a conditional dynamic panel model with general response patterns, the weak irrelevance is more stringent: d4 must be replaced by the sequential MAR $r_{i,t} \perp \mathbf{y}_{i,1}^{m,t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,t}, \mathbf{x}_{i,1}^{t}$ and the following additional condition is required:

d5. $x_{i,t} \perp \mathbf{y}_{i,1}^{m,t-1} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{o,t-1}$.

Strong ignorability for a conditional dynamic panel model requires the conditions d1-d3 and d5, and the following additional conditions:

d6. the sequential MCAR

$r_{i,t} \perp \mathbf{y}_{i,1}^{t} \big| \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t}$, and d7. $x_{i,t} \perp \mathbf{y}_{i,1}^{o,t-1} \left| \mathbf{r}_{i,1}^{t-1}, \mathbf{x}_{i,1}^{t-1} \right.$.

We emphasize that the weak and strong ignorability for the joint model, $f\left(\mathbf{y}_{i,1}^{T} \middle| \mathbf{x}_{i,1}^{T};\theta\right)$, is not equivalent to the weak and strong ignorability for the sequential model, $f(y_{i,t} \left| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1};\theta\right.)$. In the former case the ignorability requires the following conditions:

D1. two initial cuts

$f\left(\mathbf{y}_{i,1}^{T}, \mathbf{x}_{i,1}^{T}, \mathbf{r}_{i,1}^{T};\varphi\right) = f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T} \middle| \mathbf{x}_{i,1}^{T};\varphi_1\right) f\left(\mathbf{x}_{i1}^{T};\varphi_2\right)$,

$f\left(\mathbf{y}_{i,1}^{T}, \mathbf{r}_{i,1}^{T} \middle| \mathbf{x}_{i,1}^{T};\varphi_1\right) = f\left(\mathbf{y}_{i,1}^{T} \middle| \mathbf{x}_{i,1}^{T};\theta\right) f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{y}_{i,1}^{T}, \mathbf{x}_{i,1}^{T};\phi\right)$, and

D2. the MAR condition $f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{y}_{i}^{o}, \mathbf{y}_{i}^{m}, \mathbf{x}_{i,1}^{T};\phi\right) = f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{y}_{i}^{o}, \mathbf{x}_{i,1}^{T};\phi\right)$ to ensure weak ignorability, or

D3. the MCAR condition $f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{y}_{i}^{o}, \mathbf{y}_{i}^{m}, \mathbf{x}_{i,1}^{T};\phi\right) = f\left(\mathbf{r}_{i,1}^{T} \middle| \mathbf{x}_{i,1}^{T};\phi\right)$ to ensure strong ignorability.

The equivalence between the ignorability defined for the joint model and for the sequential model is true only if x is strongly exogenous for the parameters of the dynamic model of interest. We use the definition of strong exogeneity introduced by Engle *et al.* (1983); that is, (y,r) does not Granger cause x, and x is weakly exogenous for the parameter of interest. Therefore, the strong exogeneity of x includes the condition d1, d5 and d7.

We remark that if the model, $f(y_{i,t} \left| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1};\theta\right.)$, is used to forecast y given the value of x, then we need the strong exogeneity of x. For example, this is the case in causal inference, when the counterfactual response $y_{i,t}^{m}$ is forecasted conditioning on $(\mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1})$ to assess the average effect of a treatment. In this case, $r_{i,t}$ is equal to *1* if a person is treated in the time period t, and *0* otherwise. In causal inference, we should be aware that any conditioning variable, x, should be strongly exogenous. In other words, the Granger non-causality condition,

$$f(x_{i,t} \left| \mathbf{x}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1}\right.) = f(x_{i,t} \left| \mathbf{x}_{i,1}^{t-1}\right.),$$

must be satisfied.

2.8 The MAR condition according to Robins et al.

Robins and several different co-authors (Robins, Rotnitzky and Zhao 1995, Gill and Robins 1997, Robins and Gill 1997) have given definitions of MAR and MCAR for multivariate data in papers. In this section, we present these definitions and outline their differences from ours.

The definition of MAR for monotone response patterns in Robins and Gill (1997) and Robins, Rotnitzky and Zhao (1995) are both equivalent to the sequential MAR definition given in Section 2.6 for the attrition case, $r_{i,t} \perp y_{i,t} \left| \mathbf{r}_{i,1}^{t-1}, \mathbf{y}_{i,1}^{t-1}\right.$. The *k-sequential coarsening at random* (denoted briefly by '*k*-sequential CAR') definition, given by Gill and Robins (1997) and adapted for the attrition case, is again equal to the sequential MAR. In Appendix C, we prove this claim and we present the definitions of a *k*-sequential coarsening and of *k*-sequential CAR given by Gill and Robins (1997).

We remark that these definitions are not sufficient to ensure a correct likelihood-based inference on the parameters of the conditional model, $f\left(y_{i,t} \middle| \mathbf{y}_{i,1}^{t-1};\theta\right)$. Two additional conditions are necessary: the sequential cut (a1) and the Granger non-causality (a2), $y_{i,t} \perp \mathbf{r}_{i,1}^{t-1} \left| \mathbf{y}_{i,1}^{t-1}\right.$.

Moreover, we emphasize that the above MAR conditions defined for the sequential model $f\left(r_{i,t}\middle|\mathbf{y}_{i,1}^{t},\mathbf{r}_{i,1}^{t-1}\right)$, which we call sequential MAR conditions, and the MAR condition for the multivariate model $f\left(\mathbf{r}_{i,1}^{T}\middle|\mathbf{y}_{i,1}^{T}\right)$, are not equivalent. As a matter of fact, Robins and Gill find examples in which the sequential MAR condition does not ensure the MAR one. In borrowing from model reduction theory, it is possible to define conditions such that the sequential MAR condition is equivalent to the MAR condition for the joint model defined for T consecutive periods. What is missing in the work of Robins *et al.* is that the MAR condition is not enough to ensure the weak ignorability condition; indeed, the initial cut in 1' must also be satisfied. In terms of conditions on the sequential models, the initial cut is satisfied if and only if the sequential cut (a1) and the Granger non-causality (a2) are satisfied (see Engle *et al.* 1983). This is the reason why the sequential MAR definition does not ensure the MAR condition in any situation. Model reduction theory allows us to prove that when the initial cut in 1' is satisfied (or the sequential cut in a1) and the Granger non-causality in a2 are satisfied, then the sequential MAR and the MAR concepts are equivalent.

When the response pattern is not monotone, following the suggestion given in Robins, Rotnitzky and Zhao (1995), we can decide to make inference using only the sub-vector of consecutive observed variables and discharge all the observations after the first non-response. So, for example, if $\mathbf{r}_{i,1}^{7}=(1,1,1,0,1,0,1)$, then we use only the observations on the variable of interest, say y, for the first 3 waves. Let $s_{i,t}=I(\mathbf{r}_{i,1}^{t}=\mathbf{1})$, where $I(.)$ is a dummy variable, taking value 1 if the event between brackets is true and 0 otherwise; then we can artificially assume that $y_{i,t}$ is observed when $s_{i,t}=1$, and missing otherwise. In this way, the response pattern is artificially monotone and the above definition of sequential MAR applies. As remarked by Robins, Rotnitzky and Zhao (1995), this a good expedient that allows us to make a correct likelihood-inference based on the sub-sample of monotone response patterns when $s_{i,t}\perp y_{i,t}\middle|\mathbf{s}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t-1}$. In any case, this method does not use all the information available, and is therefore inefficient.

Robins, Rotnitzky and Zhao (1995) show that if we want to use all the information, we should impose an additional condition to ensure MAR. This additional condition is:

$$\Pr\left(r_{i,t}=0\middle|y_{i,1}^{o,t-1},y_{i,1}^{m,t-1},r_{i,1}^{t-1},y_{i,t+1}^{T}\right)=\Pr\left(r_{i,t}=0\middle|y_{i,1}^{o,t-1},r_{i,1}^{t-1}\right).$$

We emphasize that the above additional condition can be rewritten as the following two conditions:

(1) $r_{i,t}\perp\mathbf{y}_{i,t+1}^{T}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t-1}$, or equivalently, $y_{i,t}\perp\mathbf{r}_{i,1}^{t-1}\middle|\mathbf{y}_{i,1}^{t-1}$;

(2) $r_{i,t}\perp\mathbf{y}_{i,1}^{m,t-1}\middle|\mathbf{y}_{i,1}^{o,t-1},\mathbf{r}_{i,1}^{t-1}$.

Condition (1) is the Granger non-causality condition (a2 in Section 2.5), which is a necessary condition to ensure weak ignorability, even in the case of monotone response. Condition (2), together with $r_{i,t}\perp y_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t-1}$, is equivalent to the sequential MAR condition given in Section 2.5.

As proved in Appendix C, in the case of the non-monotone response pattern, the k-sequential CAR in Gill and Robins (1997) is different from both our definition of MAR and the one given by Robins, Rotnitzky and Zhao (1995).

In conclusion, the definitions of sequential MAR are not equivalent in the different papers of Robins and co-authors. Borrowing from reduction model theory we have clarified what is missing in the definitions by Robins et al. for the special case of panel data, i.e. for the case in which there is a sequential order for the observations on the same units.

2.9 Further extensions of the MAR and MCAR conditions

The concepts of Granger causality, sequential cut, and strong and weak exogeneity are meaningful when working with time series analysis. In the previous sections, we have shown that these concepts are very useful for panel data too, which can be viewed as a set of time series. In particular, we have shown their usefulness in extending the definitions of MAR and MCAR from cross-sectional data to panel data. By analogy, the same extension applies to the definitions of coarsening at random given in Heitjan and Rubin (1991) and described in Appendix A.

The same type of extension can be useful in causal inference when the treatments or risk exposures, the effects of which are to be evaluated, are time varying. In particular, this extension is helpful in disentangling some of the misunderstandings between Holland and Granger (see Holland 1986). Holland's (1986) attempt to use the definition of Granger causality in causal inference is misleading because he considers the evaluation of the effect of a treatment lasting in a single period. Granger causality is only meaningful when there are repeated observations across time and when attention is focused on a sequential model conditioning to past information, (see Granger 1986). I agree instead with Holland (1986) when, in his reply to Granger, he explains how the application of Rubin's model is not limited to cross-sectional data but may be extended to situations in which there are time series data for each unit or the so-called panel or longitudinal data.

As Holland (1986) remarks, in the 1980s, there were no applications of causal inference to longitudinal data, but now there are numerous examples of such studies (see, for example Robins, Greenland and Hu 1999). In these applications, the Granger causality concept is useful to help understand which conditions are necessary to make a correct causal inference and to clarify the difference between the causal concepts developed by Granger and Rubin.

3. Limits of Some Tests for MAR and MCAR in Longitudinal Data

Both the MAR and MCAR conditions require that the selection mechanism does not depend on unobserved variables. Clearly it is hard to verify dependence on unobserved variables whose values are unknown. Tests for the MAR or the MCAR conditions that verify restrictions on the parameters of the model of interest ignoring the selection mechanism, or, vice versa, on the parameters of the selection mechanism disregarding the model of interest, fail the objective, at least partially.

In this section, we outline the limitations of the procedures proposed by Little (1988), Diggle (1989), Park and Davis (1993) and Taris (1996, 1997) in detecting the selection problem. These procedures are only able to detect the MCAR conditions in part, and they cannot check the MAR assumption. These procedures investigate the dependence of the selection mechanism on the observed variables, but they cannot control for the selectivity caused by the dependence of the selection mechanism on missing variables.

3.1 Limits of the Little and Park-Davis tests

The Little (1988) and Park and Davis (1993) tests are based on a common idea: to divide units into groups according to the missing (response) pattern, $(\mathbf{r}_{i,1}^{T})$,[6] and to estimate the model of interest for each group separately, then to test the MCAR condition by verifying if the estimated

[6] For example, for a panel of T waves there are 2^T possible response patterns and therefore 2^T corresponding groups in which a unit may belong.

parameters of the models, associated with each missing pattern, are different. Little considers the normal probability distribution for a continuous variable, *y*, subjected to non-response, and tests the MCAR assumption by a likelihood ratio test. Park and Davis consider the distribution of a discrete variable, *y*, conditional on a set of explanatory variables, and use a Wald test, instead of a likelihood ratio test, to verify the MCAR. Both tests verify a condition that is only necessary but not sufficient to guarantee the MCAR assumption. Suppose that *T* different repeated values are observed for the unit, *i*, for the variable, *y*, $\mathbf{y}_{i,1}^{T}$, then the Little test verifies if $\mathbf{y}_{i,1}^{t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{r}_{i,1}^{t.}, \mathbf{y}_{i,1}^{(o)t-1}$, while the Park and Davis test verifies if $\mathbf{y}_{i,1}^{t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{r}_{i,1}^{t.}, \mathbf{y}_{i,1}^{(o)t-1}, \mathbf{x}_{i.1}^{t}$, where $\mathbf{x}_{i.1}^{t}$ are variables that are always observed.

The null hypothesis used in both tests is inadequate. The reason for this inadequacy is more evident when the missing data problem is limited to the attrition problem. Let *y* be a variable that we observe on *N* units repeatedly in time, up to the drop out of the unit from the panel or up to *T*, the last wave of the panel. Little (1988) assumes that, under MCAR, $\mathbf{y}_{i,1}^{T}$ is distributed as $N(\mu,\Sigma)$, no matter what the response pattern, $\mathbf{r}_{i,1}^{T}$, is. Then, Little (1988) tests MCAR verifying if the sub-vector of the observed variables is distributed as a multivariate normal with mean equal to the corresponding sub-vector of μ and sub-matrix of Σ, of the multivariate normal distribution for $\mathbf{y}_{i,1}^{T}$. In the case of attrition, the sub-vector of observed variables for a generic unit dropping out after *t* periods is $\mathbf{y}_{i,1}^{t}$ and we denote with $\mu^{(t)}$ and $\Sigma^{(t)}$ the mean vector and the variance matrix corresponding to the sub-vector of first *t* elements of μ, and to the *t*x*t* principal sub-matrix of Σ. Let m_t be the number of units that drop out of the panel at period (t+1), let $\overline{\mathbf{y}}^{(t)} = \frac{1}{m_t}\sum_{j=1}^{m_t}\mathbf{y}_{j,1}^{t}$, and let $\hat{\mu}^{(t)}$ be equal to the sub-vector of the first *t* elements of the maximum likelihood estimator of μ, then the Little test statistic equals $T_L = \sum_{t=1}^{T} m_t(\overline{\mathbf{y}}^{(t)} - \hat{\mu}^{(t)})'\Sigma^{(t)-1}(\overline{\mathbf{y}}^{(t)} - \hat{\mu}^{(t)})$. Little asserts that under the MCAR assumption, T_L is distributed as a Chi-square, with $\frac{T \times (T\text{-}1)}{2}$ degrees of freedom. This assertion is true; however, the same distribution remains valid under the weaker assumption that $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t}$.

Little's test cannot verify if an observable variable, $y_{i,t}$, given its past values, is independent from $\mathbf{r}_{i,1}^{t}$; in fact, if $y_{i,t}$ is observable, $\mathbf{r}_{i,1}^{t}$ is always equal to the vector of ones. In other words, Little's test cannot verify the MAR condition, $y_{i,t} \perp \mathbf{r}_{i,1}^{t} \big| \mathbf{y}_{i,1}^{t-1}$, but can only check the condition, $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t}$. We can prove that the last condition is equivalent to the hypothesis that *y* does not Granger cause *r*, $r_{it} \perp \mathbf{y}_{i,1}^{t-1} \big| \mathbf{r}_{i,1}^{t-1}$.[7] In conclusion, the Little test verifies a condition that is necessary but not sufficient for MCAR, and that is neither necessary nor sufficient for the MAR assumption (see Section 2).

The same comments apply to the Park and Davis test, if we change the above conditional independence hypothesis by adding a set of explanatory variables, *x*, among the conditioning variables, and consider a discrete distribution for the variable *y*.

[7] For a formal proof, see Florens and Mouchart (1982).

An equivalent reasoning is valid when the missing problem is more general than the attrition problem. The true null hypothesis of the Little test is $\mathbf{y}_{i,1}^{t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{r}_{i,1}^{t}, \mathbf{y}_{i,1}^{(o)t-1}$ or equivalently $r_{i,t} \perp \mathbf{y}_{i,1}^{(o)t-1} \big| \mathbf{r}_{i,1}^{t-1}$; again, this is a condition that is necessary but not sufficient for MCAR.

3.2 Limits of the Diggle test

Diggle (1989) has proposed a class of tests to verify if the attrition in a panel survey occurs at random. Given a panel with T waves, the units can be observed for a number of consecutive periods ranging from *1* to T. The tests proposed by Diggle verify if units that dropout at the (*t+1*)*-th* wave represent a random sample of units that drop out after the (*t+1*) or more waves. He introduces a score function of the observed past variables $\mathbf{y}_{i,1}^{t}$, $h(\mathbf{y}_{i,1}^{t})$, that should be linked to the probability of drop out, and tests if the score functions for the units dropping out after (*t+1*) times are a random sample from the set of scores for units that drop out in the (*t+1*)*th* wave or later. A possible test used to verify this is a Kolmogorov-Smirnov statistic test.

In other words, Diggle (1989) verifies whether the distribution of $\left\{\mathrm{h}(\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}) \mid \mathbf{r}_{i,1}^{t} = 1, r_{i,t+1} = 1\right\}$ is equal to the distribution of $\left\{\mathrm{h}(\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}) \mid \mathbf{r}_{i,1}^{t} = 1, r_{i,t+1} = 0\right\}$; that is, whether the condition $\left\{\mathrm{h}(\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}) \perp r_{i,t+1} \mid \mathbf{r}_{i,1}^{t} = 1\right\}$ holds. Let us assume that the function h is such that $\left\{\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}} \perp r_{i,t+1} \mid \mathbf{r}_{i,1}^{t} = 1, \mathrm{h}(\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}})\right\}$; that is, h is, given the past information of r, a balancing score, as defined by Rosenbaum and Rubin (1983). In this case, testing $\left\{\mathrm{h}(\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}) \perp r_{i,t+1} \mid \mathbf{r}_{i,1}^{t} = 1\right\}$ is equivalent to testing $\left\{\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}} \perp r_{i,t+1} \mid \mathbf{r}_{i,1}^{t} = 1\right\}$; that is the condition that y does not Granger cause r, which is not the MAR condition.

Diggle suggests choosing a function h that reflects the probability that $r_{i,t+1}$ =1 as a function of $\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}$; that is, he implicitly suggests using the propensity score, $\Pr(r_{i,t+1} = 1 \mid \mathbf{y}_{\mathrm{i},1}^{\mathrm{t}}, \mathbf{r}_{i,1}^{t} = 1)$. As proven by Rosenbaum and Rubin (1983), the propensity score is the coarsest balancing score; in other words, any other balancing score is a function of the propensity score.

In conclusion, the Diggle test verifies the Granger non-causality condition, $\left\{\mathbf{y}_{\mathrm{i},1}^{\mathrm{t}} \perp r_{i,t+1} \mid \mathbf{r}_{i,1}^{t} = 1\right\}$. However, it is not able to verify if $\left\{r_{i,t+1} \perp y_{i,t+1} \big| \mathbf{r}_{i,1}^{t} = \mathbf{1}, \mathbf{y}_{i,1}^{t}\right\}$, and so it is not a test for the MAR or, as defined by Diggle, for random dropouts.

3.3 Limits of the Taris test

Let $\Pr(r_{i,t+1} = 0 \mid \mathbf{r}_{i,1}^{t} = 1, \ \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t})$ be the probability to drop out at a specific wave, t, for a generic unit, i, conditioning on its permanence in the panel until wave (*t-1*) and on a set of explanatory variables. Let τ be the time of permanence of a unit in the panel; then we can rewrite the above probability as:

$$\Pr(\tau_i = t \mid \tau_i > t\text{-}1, \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t}),$$

which is a discrete hazard function. If the data are MCAR, then the hazard function should depend neither on observed variables nor on unobserved ones, and should be constant across waves; that is:

$$\Pr(\tau_i = t \mid \tau_i > t\text{-}1, \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t}) = c.$$

A very interesting result for the hazard function is stated by the "lemma" of movers and stayers, which states that when the distribution of a duration T, conditional on a set of variables x, is exponential with a parameter $\lambda(x)$, a function of x, and x follows any distribution for which the first derivative exists, then the non-conditional hazard function of *T*, marginalized with respect to x, is time decreasing (see Lancaster 1990). This means that a negative time dependence of the hazard function may be caused by the omission of relevant explanatory variables. Therefore it is necessary to distinguish between spurious and true time dependence.

Under the assumption that there is no true time dependence, a decreasing hazard function implies that data are not MCAR, while a constant hazard implies that we would not reject the MCAR condition.

This is the idea developed by Taris (1996, 1997), who says that '…a decreasing non-response for every successive wave indicates that non-response is selective to a degree.' Taris's idea is very useful to verify the MCAR condition. Taris also explains that it is possible to control for observed variables by trying to identify different groups of the population for which the hazard function is constant. In this case we would say that data are MAR but not MCAR. Taris does not use the conditional duration model approach in which variables enter as explanatories; rather he uses the Markov chains approach (the simple first order Markov chain, the mixed Markov chain and the mover-stayer model).

We think that the conditional duration model approach can be useful to detect the MAR condition. A conditional duration model is more general then a Markov chain model because it allows for time non-homogeneity, and it may be very useful in distinguishing between observed and unobserved heterogeneity causing the spurious time dependence.

If, after controlling for all observed variables in the hazard model, there is still a time dependence, then we should conclude that the data are neither MAR nor MCAR; whereas in the absence of time dependence, we cannot reject that data are MAR. If, without controlling for any explanatory variables, there is time independence, then we cannot reject the MCAR assumption.

Obviously we should not exclude a priori the assumption that the hazard function may be the result of a mixture of different hazard functions for different populations, as in the mixed Markov chain.

In conclusion, the Taris idea of verifying the MCAR and MAR conditions by checking the time dependence is very useful, but its validity is based on the assumption that the hazard function has no true time dependence. This assumption may not be true. Indeed, there may be a conditioning problem in the behavior of the person. For example, if a person is always contacted by the same interviewer, it may be that the propensity to drop out decreases from one wave to another. Furthermore, in testing the MAR condition, a misspecification of the selection mechanism can distort the results.

3.4 Limits of the variable addition test

Another type of test that has been suggested to verify the relevance of the selection mechanism is the variable addition test. This is a simple test that verifies the influence of variables associated with the non-response patterns on the regression model of interest. These variables are added to the regression model of interest as explanatory variables. If these added variables are not significant, then the selection mechanism is considered ignorable.

One should be careful in choosing the additional variables. In the case of the attrition problem, it is useless to add $r_{i,t-1}$ to a regression equation at the time *t* containing also a constant; in fact, $r_{i,t-1}$ always takes value *1*. If there are time effects in the regression, it is also inappropriate to use $\sum_{t=1}^{T} r_{i,t}$.

The MAR condition $y_{i,t} \perp \mathbf{r}_{i,1}^{t} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$ is impossible to verify, because we only have information on $y_{i,t}$ when $\mathbf{r}_{i,1}^{t} = \mathbf{1}$. We are only able to verify if $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1} = \mathbf{1}$, that is, if $r_{i,t} \perp \mathbf{y}_{i,1}^{t-1} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{t-1} = \mathbf{1}$, which is not sufficient to ensure the MCAR and MAR conditions.

Verbeek and Nijman (1992) presented the results of a Monte-Carlo analysis of the properties of the variable addition tests and found that in some cases, the variable addition tests have no power. In particular, when they used the following model of interest and missing data mechanism for the simulation experiment:

$$y_{i,t} = x_{i,t}\beta + \alpha_i + \varepsilon_{i,t}\,, \tag{1}$$

$$\Pr(r_{i,t} = 1) = \Pr(r_{i,t}^{*} > 0) = \Pr(\gamma_0 + \gamma_1 x_{i,t} + \xi_i + \eta_{i,t} > 0)\,, \tag{2}$$

where $\varepsilon_{i,t}$ and $\eta_{i,t}$ are error terms i.i.d. with mean zero, $V(\varepsilon_{i,t}) = \sigma_\varepsilon^2$, $\mathrm{V}(\eta_{i,t}) = \sigma_\eta^2$ and $Cov(\varepsilon_{i,t}, \eta_{i,t}) = \sigma_{\varepsilon,\eta}$; α_i and ξ_i are random effects i.i.d. with mean zeros, $V(\alpha_i) = \sigma_\alpha^2$, $\mathrm{V}(\xi_i) = \sigma_\xi^2$, $Cov(\alpha_i, \xi_i) = \sigma_{\alpha,\xi}$ and $\sigma_\xi^2 + \sigma_\eta^2 = 1$; then, they found that each of the following variables, $\sum_{t=1}^{T} r_{i,t}$, $\prod_{t=1}^{T} r_{i,t}$, $r_{i,t-1}$, added to equation (1) were not significant.

In the following we prove that the additional variable tests proposed by Verbeek and Njiman (1992) are adequate to check departure from MAR caused by a correlation between the random effects in the two equations, while they are not adequate to check departure caused by the correlation between the error terms. Since Verbeek and Nijman (1992) do not allow for a severe selection bias caused by the correlation between random effects, the little power of the additional variable tests follows. In the reference experiment situation in Verbeek and Nijman (1992), the correlation between ξ and η is 0.5, but the importance of the random effects in both equations is too low, the ratios $\sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\varepsilon^2)$ and $\sigma_\xi^2 / (\sigma_\xi^2 + \sigma_\eta^2)$ are 0.1, so that the resulting selection bias is not severe and the power of the tests is small.

To prove that the additional variable tests proposed in Verbeek and Nijman (1992) cannot be used to verify departure from the MAR caused by correlation between error terms, we consider the case of a null correlation between the random effects in the equations (1) and (2). If the correlation between random effects is *0*, then the following independence conditions hold: $y_{i,t} \perp \mathbf{r}_{i,1}^{t-1} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$ and $r_{i,t} \perp \mathbf{y}_{i,1}^{t-1} \big| \mathbf{x}_{i,1}^{t}, \mathbf{r}_{i,1}^{t-1}$ (that is, $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$). By consequence, the equation (1) is not affected by $\mathbf{r}_{i,1}^{t-1}$ and $\mathbf{r}_{i,t+1}^{T}$, but only by $r_{i,t}$. Obviously the dependence between $y_{i,t}$ and $r_{i,t}$ cannot be verified because we observe $y_{i,t}$ only when $r_{i,t} = 1$.

The above authors have carried out the same simulation exercise for the quasi-Hausman test (a test which verifies if the model coefficients for the balanced and unbalanced panels are equal) and have found that the power is better but non-satisfactory. This is again a consequence of the fact that, ignoring the random effects because of their little importance, $y_{i,t} \perp \mathbf{r}_{i,1}^{t-1} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$ and $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$, so that

$$f\,(y_{i,t} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}, \mathbf{r}_{i,1}^{T} = 1) = f\,(y_{i,t} \big| \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}, r_{i,t} = 1)\,,$$

and the balanced and unbalanced panels give the same results.

When instead, the authors simulated the following model for the missing data mechanism:

$$\Pr(r_{i,t} = 1) = \Pr(r_{i,t}^{*} > 0) = \Pr(\gamma_0 + \pi\,\overline{x}_i + \xi_i + \eta_{i,t} > 0)\,, \tag{3}$$

the power of the variable addition tests and of the quasi-Hausman tests increased. This is not surprising since in this case, x is not strictly exogenous for the parameters of model (1) and $y_{i,t} \perp \mathbf{r}_{i,t+1}^{T} \mid \mathbf{x}_{i,1}^{t}, \mathbf{y}_{i,1}^{t-1}$. This means that variables that are linked to the future response path $\mathbf{r}_{i,t+1}^{T}$ affect the model (1). The results of Verbeek and Nijman (1992) support this claim; in fact, the power of the tests obtained by adding the variables $\sum_{t=1}^{T} r_{i,t}$ e $\prod_{t=1}^{T} r_{i,t}$ is good, while the power is very small when the variable $r_{i,t-1}$ is added.

The same type of reasoning implies that the quasi-Hausman tests are more powerful when model (3) is used for simulation instead of model (2), and the results again support our conclusion.

Finally, Verbeek and Nijman (1992) also computed the power for the Lagrange multiplier test and found that it is good in both simulations. In fact, the Lagrange multiplier test is the only one of the three tests used that correctly takes account of the joint specification of the model of interest and selection mechanism.

It seems that the simulation results obtained by Verbeek and Nijman (1992) are in support of the observation that tests trying to verify the ignorability of the selection mechanism without jointly specifying the model of interest and selection mechanism can be misleading. As these tests under-reject the null hypothesis of ignorability, their usefulness in the detection of the selection problem is questionable.

3. Conclusions

Rubin (1986) proves that in order to make correct likelihood-based inference, we need two conditions: the MAR condition and the variation-free condition for the parameters of the model of interest and the selection mechanism. In this paper, we have defined the weak ignorability of the selection mechanism as the set of conditions necessary and sufficient to make correct and efficient inference based on the likelihood function. Using the terminology of model reduction theory, we have shown that weak ignorability is satisfied if the model of interest and selection mechanism operate a statistical cut, and if the MAR condition is true. In borrowing from model reduction theory, we have extended the definitions of weak ignorability to the panel data case. Two definitions of weak ignorability may be given: one in terms of a joint model of interest, defined for T consecutive waves, and another in terms of a sequential model, corresponding to a dynamic model of interest and defined for a single time period. We have proved that weak ignorability for a joint model of interest requires a MAR condition and an initial cut, whereas weak ignorability for a dynamic model requires a sequential cut, a Granger non-causality condition and a sequential MAR condition. Moreover, we have shown that, if the model of interest is conditional on a set of explanatory variables, then some additional conditions are necessary. Substituting MAR with MCAR in the definition of weak ignorability, we have obtained the strong ignorability definition, which is the condition ensuring a correct inference for any type of inference methodology.

The extension of weak and strong ignorability to the case of dynamic panel models has allowed us to emphasize the failure of some tests proposed in the literature to verify the MAR and/or the MCAR conditions. Indeed, we have proved that the null hypothesis of some tests is given by an assumption that is not necessary for MAR and which is necessary but not sufficient for MCAR.

Furthermore, the formal definition of weak and strong ignorability has helped us to emphasize some of the limitations of the MAR and MCAR definitions given by Robins and co-authors, and to disentangle some of the misunderstandings that occurred between Holland and Granger concerning the concept of causality in the causal inference.

Appendix A: Definition of CAR following Heitjan and Rubin (1991)

Heitjan and Rubin (1991) consider a general coarsening mechanism $f(r|y;\phi)$, in which r is a variable indicating the level of coarsening. For example, if there is only a level of coarsening r, is a dummy variable and we assume that y is coarsened if $r=0$ and perfectly observable if $r=1$. When $r=0$ we have a piece of information about y that is not precise. For example, in the case of missing data, y is not observable; in grouped data, y is known to belong to a sub-space of its domain; in right censored duration data, y belong to (c,∞) where c is the censor value. In general coarsened data occur when we do not know the exact value of y, but we know that y belongs to a sub-space of $\mathbf{Y}$. Let $\tilde{y}$ be the coarse variable, which defines the sub-space to which y belong, then $\tilde{y}=y$ when $r=1$ and $\tilde{y}\in \mathbf{B}\subset \mathbf{Y}$ if $r=0$. In the case of missing data $\mathbf{B}=\mathbf{Y}^m$ and it is often equal to the entire space $\mathbf{Y}$.

More generally, r may be a continuous variable, with a sample space given by $\mathbf{R}$, that determines the coarsening mechanism, so that $\tilde{y}$ can be expressed as a function of y and the variable r, $\tilde{y}=\tilde{Y}(y,r)$. The distribution function of r given y, $f(r|y;\phi)$, is the process that determines the level of precision in measuring y. In the case of missing data the coarsening mechanism is a selection process or missing data mechanism, in the grouped data it is a grouping mechanism, in the causal inference it is an assignment process, and so on.

The definition of *coarsening at random* (CAR) given by Heitjan and Rubin (1991), that generalizes the missing at random (MAR) given by Rubin (1976), is the following one: y is coarsened at random if, for each fixed value $\tilde{y}$, $f(r|y;\phi)$ takes the same value for all $y\in \tilde{y}=\tilde{Y}(y,r)$.

The MAR definition (1) given in Section 2 is equal to the MAR given in Heitjan and Rubin (1991). In fact when y is observed, $\tilde{y}$ is not an interval but a point, so the requirement that $f(r|y;\phi)$ takes the same value for all $y\in \tilde{y}=\tilde{Y}(y,r)$ is always satisfied. Therefore the Heitjan and Rubin (1991) MAR definition reduces to require that $f(r|y;\phi)$ takes the same value for all $y^m\in \mathbf{Y}^m$, that is the definition of MAR in Section 2.

The CAR condition together to the variation free condition ensure that the censored likelihood, L_c, and the likelihood with informative missing data, L_I, are equal. Indeed the two likelihood functions are respectively given by the following expressions:

$$L_c=\int_{\tilde{y}} f(y;\theta)dy=\left(f(y^o;\theta)\right)^r\left(\int_{\mathbf{Y}^m} f(y^m;\theta)dy^m\right)^{1-r}, \qquad (1)$$

and

$$\begin{aligned} L_I&=\int_{\tilde{y}}\int f(y,r;\theta,\phi)\, f(\tilde{y}\,|\,y,r)drdy=\int_{\tilde{y}} f(y;\theta)\int f(\tilde{y}\,|\,y,r)\, f(r\,|\,y;\phi)drdy= \\ &=\int_{\tilde{y}} f(y;\theta)\, f(\tilde{y}\,|\,y;\phi)dy\ , \end{aligned} \qquad (2)$$

where the integration is respect to the underlying dominating measure, a Lebesgue measure or a counting measure, and $f(\tilde{y}\,|\,y,r)$ is the conditional degenerate distribution of $\tilde{y}$ given y and r

$$f(\tilde{y}\,|\,y,r)=\begin{cases}1 \text{ if } \tilde{y}=\tilde{Y}(y,r)\\ 0 \text{ if } \tilde{y}\neq\tilde{Y}(y,r)\end{cases}.$$

Under CAR $f(r|y;\phi)$ takes the same value for any $y\in \tilde{y}=\tilde{Y}(y,r)$ so $\int f(\tilde{y}\,|\,y,r)\,f(r\,|\,y;\phi)dr = f(\tilde{y}\,|\,y;\phi)$ is a constant, say α, for any $y\in \tilde{y}$ and we can rewrite (2) as $\left(\alpha\int_{\tilde{y}} f(y;\theta)dy\right)$, that is proportional to the likelihood (1). The proportionality between (1) and (2) under CAR ensures that inference on θ based on the censored likelihood or on the likelihood with informative missing data is equal.

Sometimes r is unknown. An example is given by the case of a survey in which some units give a rounded response and some other give the exact value, but we cannot distinguish between the two types of units. When r is unknown, the definition of coarsened at random is: y is coarsened at random if, for each fixed value $\tilde{y}$, $f(\tilde{y}\,|\,y;\phi)=\int f(\tilde{y}\,|\,y,r)\,f(r\,|\,y;\phi)dr$ takes the same value for all $y\in \tilde{y}=\tilde{Y}(y,r)$.

For a formal proof of the equivalence between inference based on likelihood (1) and (2) see Heitjan and Rubin (1991), for detailed examples see Heitjan (1993).

Appendix B: The case of a deterministically censored variable

In this section we present a very simple example of a censored variable to show that MAR condition does not require that the selection mechanism is constant for any y but only for any $y\in Y^m$.

Let y be a continuous variable with support $\mathbf{Y}=(-\infty,+\infty)$ and let us assume that we observe y only when its value is lower than or equal to a constant c, then $\mathbf{Y}^m=(c,\infty)\subset\mathbf{Y}$ and y is MAR because for any value greater than c the probability to observe y is equal to 0.

In this specific example the likelihood (1) in appendix A becomes

$$\int_{\tilde{y}} f(y;\theta)dy=\left(f(y^o;\theta)\right)^r\left(\int_{\mathbf{Y}^m} f(y^m;\theta)dy^m\right)^{1-r}=\left(f(y^o;\theta)\right)^r\left(1-F(c;\theta)\right)^{1-r}.$$

The selection mechanism $f(r|y;\phi)$ is deterministic, in fact

$$r=\begin{cases}1 \text{ if } y\le c \text{ with probabiltiy } 1\\ 0 \text{ if } y>c \text{ with probabiltiy } 1\end{cases}.$$

When y is missing $f(\tilde{y}\,|\,y^m;\phi)=\sum_{r=0}^{1} f(\mathbf{Y}^m\,|\,y^m,r)\,f(r\,|\,y^m;\phi)=1$ for any $y^m\in\mathbf{Y}^m$, when y is observable $f(\tilde{y}\,|\,y^o;\phi)=\sum_{r=0}^{1} f(y^o\,|\,y^o,r)\,f(r\,|\,y^o;\phi)$ is also equal to 1. This allows us to write the informative likelihood (1) as

$$\int_{\tilde{y}} f(y;\theta)\,f(\tilde{y}\,|\,y;\phi)dy=\left(f(y^o;\theta)\right)^r\left(1-F(c;\theta)\right)^{1-r},$$

which is equal to the likelihood with informative missing data (2).

This equality proves that the weak ignorability of selection mechanism does not require that the selection mechanism be constant for any $y\in Y$, but only for any $y\in Y^m$.

Appendix C: The sequential CAR condition in Gill and Robins (1997) and our sequential MAR condition

A variable *X* is said to be coarsened if we cannot observe its exact value, but we know the subset of the sample space to which it belongs. In other words we observe a coarse variable χ, instead of *X*, which defines the subset to which *X* belongs.

Following Gill and Robins (1997) we assume that "… χ is a coarsening of an underlying random variable *X*. We suppose that *X* takes values in a finite space *E*. Its power set (the set of all subset of *E*) is denoted by $\mathcal{E}$. So χ takes values in $\mathcal{E}\backslash\{\phi\}$ and $X\in\chi$ with probability one."

Definition of a k-sequential coarsening: (Gill and Robins 1997) "We say that the random sets $\chi_1,...,\chi_k,\chi$ with each χ_m and $\chi\in\mathcal{E}\backslash\{\phi\}$ form a *k*-sequential coarsening of a random variable *X* if for *m=0,…, k+1,* $\chi_m \subseteq \chi_{m+1}$ with probability 1 where $\chi_o \equiv \{X\}$ and $\chi_{k+1} \equiv \chi$."

Definition of a k-sequential CAR: (Gill and Robins 1997) "A *k*-sequential coarsening is a *k*-sequential CAR if, for *m=1,…, k*, the conditional distribution of χ_m given χ_{m-1} does not depend on the particular realization of χ_{m-1} except through the fact that is compatible with χ_m. In the discrete case, this means $\Pr(\chi_m = A|\chi_{m-1} = B)$ is the same for all *B* in the support of χ_{m-1} such that $B\subseteq A$."

When the coarsening is due to the attrition problem, we prove that the *k*-sequential CAR definition of Gill and Robins (1997) is equivalent to the sequential MAR definition given in this work.

Let us consider a random sample of *N* units, for each unit *i* we observe repeatedly in time a variable *y*, which takes values in the sample space **Y**, and we denote this multivariate variable $y_{i,1}^T$, where *T* is the number of repeated observations. If $y_{i,t}$ is missing, then the successive variables, $y_{i,t+1},...,y_{i,T}$, are also unknown (this is the case of the attrition problem). Each missing variable, *y*, takes value in **Y**, so that the corresponding coarse variable, $\tilde{y}$, which defines the sub-space to which *y* belongs, is equal to the entire sample space **Y**. Let $X = [y_{i,1},...,y_{i,T}] = y_{i,1}^T$; then the coarsened multivariate variable associated to a unit *i*, for which the last *k* variables are not observed, is denoted by $\chi = [y_{i,1},...,y_{i,T-k+1},\tilde{y}_{T-K+2},...,\tilde{y}_T] = [y_{i,1},...,y_{i,T-k+1},\mathbf{Y},...,\mathbf{Y}]$.

If we define $\chi_0,...,\chi_k,\chi$ in the following way:

$\chi_0 = [y_{i,1},...,y_{i,T-k+1},y_{i,T-k+2},...,y_{T-1},y_T]$,

$\chi_1 = [y_{i,1},...,y_{i,T-k+1},y_{i,T-k+2},...,y_{T-1},\tilde{y}_T] = [y_{i,1},...,y_{i,T-k+1},y_{i,T-k+2},...,y_{T-1},\mathbf{Y}]$,

…

$\chi_k = \chi = [y_{i,1},...,y_{i,T-k+1},\tilde{y}_{T-K+2},...,\tilde{y}_T] = [y_{i,1},...,y_{i,T-k+1},\mathbf{Y},...,\mathbf{Y}]$;

then $\chi_{m-1} \subseteq \chi_m$ for any *m=0,…,k* and χ can be viewed as the result of a *k*-sequential coarsening.

To prove that χ is a *k*-sequential CAR, we have to show that $\Pr(\chi_m = A|\chi_{m-1} = B)=c$, where *c* is a constant, for all *B* in the support of χ_{m-1} such that $B\subseteq A$ (see the above definition of k-sequential CAR).

If the first (*T-1*) elements of χ_1 are not equal to the corresponding observed elements of χ_0, then $\Pr(\chi_1 = A|\chi_0 = B)$=0; so that verifying

$$\Pr(\chi_1 = A|\chi_0 = B)=c$$

is equivalent to verify that

$$\Pr\left(\tilde{y}_T = \mathbf{Y}\middle|y_{i,1},\ldots,y_{i,T-1},y_{i,T},r_{i,1}^{T-1}=1\right)=c \;\forall\; y_{i,T}\in\mathbf{Y},$$

that is, using the fact that $(\tilde{y}_{i,t} = \mathbf{Y}) = (r_{i,t} = 0)$,

$$\Pr\left(r_{i,T} = 0\middle|y_{i,1},\ldots,y_{i,T-1},y_{i,T},r_{i,1}^{T-1}=\mathbf{1}\right)=\Pr\left(r_{i.T} = 0\middle|y_{i,1},\ldots,y_{i,T-1},r_{i,1}^{T-1}=\mathbf{1}\right),$$

where *r* is the dummy indicator of response.

By analogy $\Pr(\chi_m = A|\chi_{m-1} = B)=c$ for all *B* in the support of χ_{m-1} such that $B\subseteq A$ is true if and only if

$$\Pr\left(r_{i,t} = 0\middle|y_{i,1},\ldots,y_{i,t-1},y_{i,t},r_{i,1}^{t-1}=\mathbf{1},r_{i,t+1}^{T}=\mathbf{0}\right)=$$
$$\Pr\left(r_{i.t} = 0\middle|y_{i,t},\ldots,y_{i,t-1},r_{i,1}^{t-1}=\mathbf{1},r_{i,t+1}^{T}=\mathbf{0}\right)$$

where $t=T-m+1$. Since $\Pr\left(r_{i,t} = 0\middle|y_{i,1},\ldots,y_{i,t-1},y_{i,t},r_{i,1}^{t-1}=\mathbf{1},r_{i,t+1}^{T}\neq\mathbf{0}\right)=0$ in the case of attrition, we can rewrite the last equality as

$$\Pr\left(r_{i,t} = 0\middle|y_{i,1},\ldots,y_{i,t-1},y_{i,t},r_{i,1}^{t-1}=\mathbf{1}\right)=\Pr\left(r_{i.t} = 0\middle|y_{i,t},\ldots,y_{i,t-1},r_{i,1}^{t-1}=\mathbf{1}\right),$$

that is the sequential MAR condition given in Section 2.6, $r_{i,t}\perp y_{i,t}\middle|\mathbf{r}_{i,1}^{t-1},\mathbf{y}_{i,1}^{t-1}$.

If we consider a more general response pattern, possibly non-monotone, then the definition of *k*-sequential CAR given in Gill and Robins (1997) does not correspond to our definition of sequential MAR.

Indeed, the *k*-sequential CAR condition for non-monotone response patterns is equivalent to the following condition,

$$\Pr\left(r_{i,t} = 0\middle|y_{i,1}^{oT},y_{i,t},r_{i,1}^{t-1},r_{i,t+1}^{T}\right)=\Pr\left(r_{i,t} = 0\middle|y_{i,1}^{oT},r_{i,1}^{t-1},r_{i,t+1}^{T}\right);$$

while our sequential MAR definition is

$$\Pr\left(r_{i,t} = 0\middle|y_{i,1}^{ot},y_{i,t},r_{i,1}^{t-1}\right)=\Pr\left(r_{i,t} = 0\middle|y_{i,1}^{ot},r_{i,1}^{t-1}\right).$$

References

Diggle P. (1989), Testing for random dropouts in repeated measurement data, *Biometrics*, vol. 45, pp 1255-1258

Engle R.F., Hendry D.F., Richard J.-F. (1983), Exogeneity, *Econometrica*, vol. 51, n. 2, pp 277-304

Florens J.P., Mouchart M. (1982), A note on non causality, *Econometrica*, vol. 50, no. 3, pp 583-591

Gill R.D., Robins J.M. (1997), Sequential models for coarsening and missingness, in Lin D.Y., Fleming T.R., Eds., "*Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. Lecture Notes in Statistics.*", New York, Springer-Verlag, pp 295-305

Gill R.D., van der Laan M.J., Robins J.M. (1997), Coarsening at random: characterizations, conjectures, counter-examples, in Lin D.Y., Fleming T.R., Eds. "*Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. Lecture Notes in Statistics.*", New York, Springer-Verlag, pp 255-294

Granger C.W.J. (1969), Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, vol. 37, pp 424-438

Heitjan D.F., Rubin D. (1991), Ignorability and coarse data, *Annals of Statistics*, vol. 19,n. 4, pp 2244 –2253

Heitjan D.F. (1993), Ignorability and coarse data: some biomedical examples, *Biometrics*, vol. 4, pp 1099-1109

Hendry D.F. (1995), *Dynamic Econometrics*, Advanced texts in econometrics, Oxford University Press

Holland P.W. (1986), Statistics and causal inference, *Journal of American Statistical Association*, vol. 81, no. 396, pp 945-970

Lancaster T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge, MA

Little J.A. (1988), A test of missing completely at random for multivariate incomplete data, *Journal of the American Statistical Association*, vol.77, pp 237-250

Little J.A., Rubin D.B. (1987), *Statistical Analysis with Missing Data*, John Wiley and Sons, New York

Park T., Davis C.S. (1993), A test of the missing data mechanism for repeated categorical data, *Biometrics*, vol. 49, pp 631-638

Rosenbaum P., Rubin B. D. (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, vol. 70. no. 1, pp 41-55

Robins J.M. (1997), Causal inference from complex longitudinal data, in Berkane M., Ed. "*Latent Variable Modeling and Applications to Causality. Lectures Notes in Statistics*", Springer Verlag, New York, pp 69-117

Robins J.M., Gill R.D. (1997), Non-response models for the analysis of non-monotone ignorable missing data, *Statistics in Medicine*, vol. 16, pp 39-56

Robins J.M., Greenland S., Hu F.-C. (1999), Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome, *Journal of American Statistical Association*, vol. 94, no. 447, pp 687-712

Robins J.M., Rotnitzky A., Zhao L.P. (1995), Analysis of semiparametric regression models for repeated outcomes in the presence of missing data, *Journal of the American Statistical Association*, vol. 90, no. 429, pp 106-121

Rubin D.B. (1976), Inference and missing data, *Biometrika*, vol. 63, pp 581-592

Shih W.J. (1992), On informative and random dropouts in longitudinal studies, (Letter to the Editor), *Biometrics*, vol. 48, pp 971-972

Taris T.W. (1996), Modeling nonresponse in multiwave panel studies using discrete-time Markov models, *Quality & Quantity*, vol. 30, pp 189-203

Taris T.W. (1997), On selectivity of nonresponse in discrete-time multi-wave panel studies, *Quality & Quantity*, vol. 31, pp 79-94

Verbeek M., Nijman T. (1992), Testing for selectivity bias in panel data models, *International Economic Review*, vol. 33, no. 3, pp 681-704

Verbeek M., Nijman T. (1992), "Incomplete Panel and Selection Bias", in Matyas L. and P. Sevestre, Eds., "*The Econometrics of Panel Data: Handbook of Theory and Applications*", Kluwer Accademic Publishers, Dordrecht